2009 13th International Conference Information Visualisation

Search Engine Result Visualisation: challenges and opportunities

Kenneth Treharne, David M. W. Powers
Flinders University of South Australia
{kenneth.treharne, david.powers}@flinders.edu.au

Abstract

The rank-ordered list format is the prevailing result presentation format of all the major Internet search engine providers. The format itself has seen few changes over the course of the Internet's history aside from the emergence of the snippet. However, there are indications that the result presentation format is undergoing some evolutionary changes. The discussion will examine primarily the structural aspects of alternative formats to the traditional list format and motivate a program of comparative evaluation. A summary of current structure templates provides a resource for designers of future result visualisations.

Introduction

Visualisation based approaches to search result presentation are growing in prevalence across the Internet. Evaluation must continue to show how these approaches minimise the deficiencies and maximise cognitive efficiencies of the de-facto standard ranked-ordered list. Toward this goal, we are attempting to scrutinize core components that all visualisations share to some degree; namely, the spatialization of documents and the representation of relationships between those documents.

The rank-ordered list has serviced a myriad diversity of information need. However, the generic rank-ordered list is increasingly inadequate for navigation of large quantities of information. There is evidence of some serious attempts at producing viable alternatives to the list format [8]. The Internet hosts a number of examples that propose cognitive amplification for specific stages [19] of the information retrieval process; for example, query formulation and initiation of search [39], review of results[20, 21, 23], refinement of result set [20], and use of results [20, 24].

Spatial arrangement or position is a salient feature for the representation of document relatedness. Although relatedness is a subjective matter, a relationship may be made explicit (e.g. an edge between nodes), or implicit (e.g. based on point density). Current knowledge regarding the merits of both approaches, especially in the context of search result visualisation, is imprecise. Comprehensive evaluation of the rudimentary spatial component will expand on this knowledge. In pursuit of this, we have collated examples of spatialization paradigms to aide evaluation and comparison of past, present and future result visualisation. Our approach harmonises with the work of Morse et al [29] who also take a reduced form approach for the evaluation and comparison of multiple visualisation techniques.

A viable successor to the rank-ordered list is more a question of the fundamental theoretical challenges than technological. It is likely that the infrastructure required to efficiently present search results within internet browser environments already exists in some toolkit or web service form. We have been investigating these toolkits and web services to build and evaluate experimental interfaces and in particular, using a Mozilla Firefox extension to deliver experiments over the Internet.

The next section canvases some of the advantages and disadvantages of the ranked-ordered list format. The subsequent section then introduces our initial efforts to capture different spatialization paradigms for search result presentation. The final section outlines a browser-based tool that we anticipate using for a comparative evaluation of several different reduced-form search result visualisations. Our browser-based tool will serve our research program as well as provide a useful information retrieval tool for participants even after the end of the experiment.

1. Rank-ordered result presentation

A rank-ordered list has several advantages. Its format is lean, ubiquitous and scalable; consistent, simple and intrinsic; and user and task inclusive. Such characteristics are easily measured and so early optimisations have realised. However, the rank-ordered list at present cannot support efficient and thorough information search given the high prevalence of irrelevant and unrelated results – due largely to an ongoing lack of machine understanding of human language.

The polysemous nature of language ensures that all linguistic dialogue between human and machine is ambiguous and that the result set contains many unrelated or partially related documents. However, there is no indication of relationships within and between rank-ordered results [6]; in reality, the searcher employs cognition to make judgements ad-hoc and even then,

978-0-7695-3733-7/09 $25.00 © 2009 IEEE
DOI 10.1109/IV.2009.50

comparing the rank of two unrelated documents is pointless [40]. Additionally, analysis of user behaviour reveals that many results seldom receive equal portions of attention [1, 10, 11]. A searcher would rather modify their query than look further down the result list.

The rank-ordered list is chiefly static in nature and this may disadvantage information search and exploration. A dynamic and interactive interface opens a new channel of dialogue between human and machine, thus potentially amplifying and externalising cognition to the benefit of decision-making and acquisition of information. Examples include sorting, filtering and provision of feedback and ratings that a search engine may act on autonomously. Regarding the latter, there is an optimisation necessary between the degree of interactivity offered and the degree of automation [25]. Nevertheless, a richer interaction dialogue will likely offer greater insight into the knowledge acquisition process. Thus investigating alternative result interfaces will produce both useful tools and new understanding of human cognition.

In order to reach mainstream, open-domain search, alternatives to the rank-ordered list must possess lean, generic, consistent and inclusive characteristics as well as a greater propensity to benefit human cognition during search. As new alternatives emerge [e.g. 20, 47, 49] it is practical to make comparative evaluation of contending approaches and ascertain the core components that enhance cognition during search.

2. Visualisation based result presentation

A diverse collection of visualisation-based search tools is available [8] though any associated evaluation seldom succeeds in elucidating the merits of their basal properties including spatial layout and labelling. Figure 1 illustrates our initial efforts to capture the diversity of these search tools and motivates a program of evaluation. We deliberately omit non-essential aesthetics to facilitate comparison of spatial layout. It abstracts away from the inessential and is therefore proposed as a reference for designers and in particular, those who wish to carry out comparative evaluation. As we have removed aesthetic debris, figure one emphasises primarily visualisation of inter-document relationships rather than document attributes [42].

Spatialization has two interdependent characteristics; the first one is semantic and specifies the basis for inter-document relationships, the second is spatial and specifies how best to represent those relationships visually [28]. Semantic relationships depend on analysis of a 'sampling unit' such as Metadata or thematic content [31], whereas spatial relationships depend on visual perceptions of correspondence, differentiation, connectivity and arrangement [43]. These are important to consider because high semantic coherence and spatial grouping influence the users' cognitive processing of the visualisation [30]; although, perceptual interpretation is frequently subjective.

The mutual dependence of semantic and spatial also raises issues of dimensionality, use of metaphor and labelling of documents and labelling of the spatial substrate. Information search typically deals with highly dimensional documents – words, topics, themes – and so analytical techniques [see 31, 45] are used to extract the most representative dimensions. A critical aspect of this process is the degree of information loss during dimension reduction [45]. Subsequently, the number of dimensions visualised becomes important. We have deliberately chosen to focus on two-dimensional visualisation as evidence suggests that two dimensions are better than or as good as three dimensions [41, 46]. However, omission of three-dimensional visualisation undervalues the use of metaphor that may benefit the navigation of information space. For instance, a cityscape metaphor in which buildings represent documents elucidates concepts such as relatedness (proximity) and clusters of related documents (neighbourhoods) [40].

Subjective interpretation of labels and aesthetics similarly highlights the mutual dependence of semantic and spatial. Labels can synopsise the content of individual documents and document clusters. Labels also define the structure and context of the domain of search and this is beneficial for first time searchers [7]. However, a compromise is necessary between aesthetics and functionality and remains a challenge shared by the wider visualisation community [32, 33]. Appropriate length, clarity, content and expressiveness impact on aesthetics including the degree of occlusion, readability, unambiguous referral [32], distribution, consistency [33] and local and global clutter [15]. User behaviour further confounds the need for compromise. Whilst users use 2-3 words to search for documents [16] they typically use 2-3 times that number to describe documents [38]; this exemplifies one such trade-off necessary between the descriptive power of labels and aesthetics.

In contrast, users have criticised interfaces that do not show immediately accessible result surrogates [34] so showing no labels at all is less than ideal. Presenting several results at once on demand [32] eliminates the need for concomitant number of mouse clicks, which although initially light work, can become illusorily tedious and repetitive with time. Moreover, presenting only one result at a time in a fixed screen location [20, 22, 35] provokes discontinuity and makes difficult the task of comparing documents.

Finally, the last row of figure one, added for completeness, illustrates how labelling need not be textual. Several approaches utilise iconic [36, 37] and chart-based [57, 18] representation for the labelling of semantic content within documents, while thumbnail previews facilitate comparison and recall of documents based on visual signatures [22, 47].

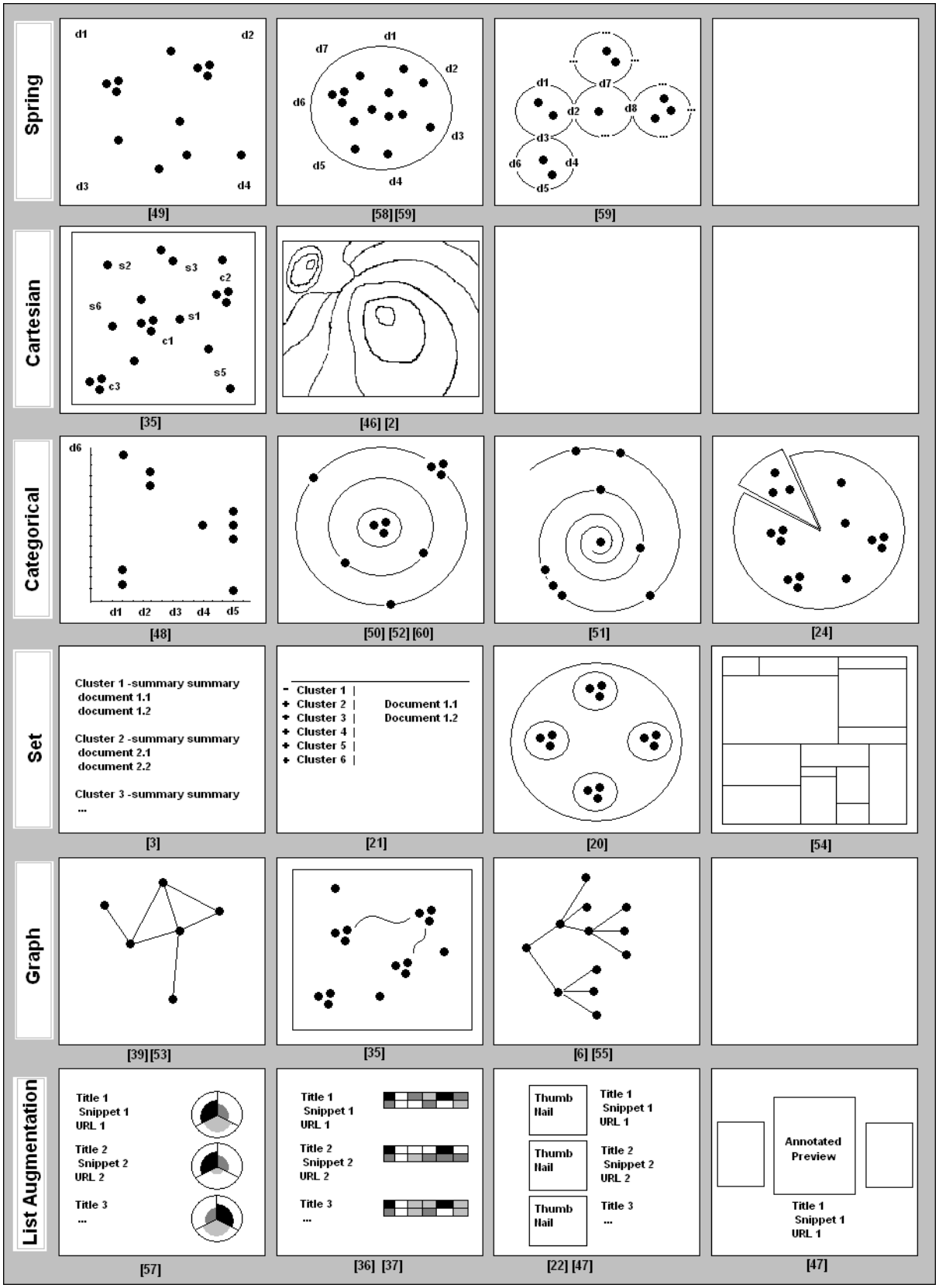

Figure 1 – Reduced form spatialization techniques for search result visualisation

3. Browser extension as evaluation tool

The potential scope for comparative evaluation of search result visualisation is quite large. In preparation for comparative evaluation, we are investigating browser-based technologies and web services as a way to deliver experiments and collect data. Online, web-based and out-of-the-laboratory experimentation is increasingly common and has a number of proponents [12, 13, 14]. There are many benefits to online delivery of experiments [12, 14, 17] including cost savings and broad demography.

Our evaluation tool is encapsulated within a Mozilla Firefox browser Extension [26]. To install the extension the participant downloads a XPI 'zippee' archive, containing program and interface script. They open this file using Mozilla Firefox, which then automates the install process. Uninstalling the extension is a simple one-step process initiated within the Add-ons Manager of Firefox.

Central to our evaluation tool are the interactive visualisations. There exist a number of browser-based technologies to produce graphical visualisations that run within or draw directly on a web page [4, 5, 9]. Our extension utilises canvas HTML5 element [27] and JavaScript for drawing and interactivity. The use of these generic technologies ensures that the evaluation is highly portable. In an earlier investigation utilising Java Applets for visualisation capability, we noticed a high degree of experiment drop out due to versioning issues with incorrect Java Runtime Environment.

With sufficient parameterisation, our evaluation extension can provide a useful information retrieval tool for participants following the evaluation. This is mutually beneficial in that we reward participants with a usable search tool, whilst they continue to provide a stream of interactive information retrieval data. Furthermore, added functionality is possible by passing search results obtained from Google, Yahoo or custom search services [e.g. 56] to a Java program running on the Java Runtime Environment. This allows the possibility to perform more sophisticated processing including custom clustering, user-feedback analysis, and maintenance of search histories. As a result, we envision this tool as a basis for an integrated suite of knowledge management tools and over time envision that users can download and plug-in new visualisations that integrate with our tool.

Figure 2 illustrates the general architecture of the current extension and its extended functionalities. To initiate a search the user types a query into the extension's browser-toolbar search box and executes the search. The extension logs the query and requests results from the source. The extension downloads results and passes them to a Java program for linguistic processing and transformation. The result interface opens as a new tab and the user is free to interact with results in a number of formats including sorted list, clustered list, or one of several visualisations. As the searcher interacts with the interface, the extension reports anonymous usage statistics. Of particular interest will be the relative portions of time spent using specific visualisation formats, query construction, search tasks, use of relevance feedback etc.

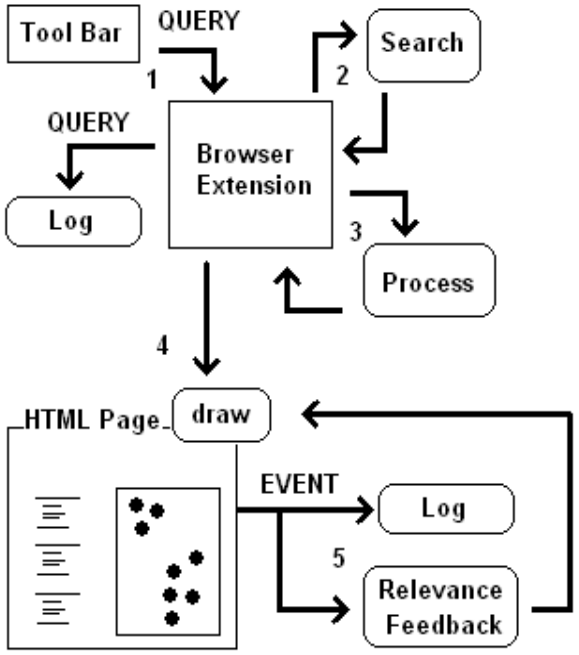

Figure 2 - Browser extension architecture.

Conclusions

In the minimum, any viable alternative to the rank-ordered list format for web search results will need to address everything wrong with the list format; moreover, it must achieve everything that the list format can do exceedingly well. If it is possible to compete with the generic result list format, it is highly possible that most of the technical infrastructure already exists to do so.

Ongoing evaluation must continue to compare the various spatialization approaches (and their textual annotation) to search result visualisation. Our collation of reduced form examples is proposed to facilitate a program of comparison and evaluation. In preparation for such an evaluation, we have been investigating the use of an internet browser extension as a delivery mechanism. Our evaluation extension uses browser-based technologies capable of displaying interactive visualisations. We believe our approach is mutually beneficial, as the extension will provide a useful information search tool for participants while delivering an ongoing source of interactive information retrieval data.

References

[1] Granka, L. A., Joachims, T. and Gay, G. 2004. Eye-tracking Analysis of User Behaviour in WWW Search, Special Interest Group on Information Retrieval, SIGIR'04, pp 478-479.

[2] Granitzer, M., Kienreich, W., Sabol, V. and Dösinger, G. 2003. WebRat: Supporting Agile Knowledge Retrieval Through Dynamic, Incremental Clustering and Automatic Labelling of Web Search Result Sets, 12th IEEE International Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises, pp 296-301.

[3] Hearst, M. and Pedersen, J. 1996. Re-examining the Cluster Hypothesis Scatter/Gather on Retrieval Results, In Proceedings of the 19th Annual International ACM Special Interest Group on Information Retrieval: Research and Development in Information Retrieval, pp 76-84.

[4] Holmberg, N., Wünsche, B. and Tempero, E. 2006. A Framework for Interactive Web-based Visualisation, Proceedings of 7th Australasian User Interface Conference, vol 50, in ACM International Conference Proceedings Series, vol 169, pp 137-144.

[5] Johnson, D.W. and Jankun-Kelly, T.J. 2008. A Scalability Study of Web-Native Information Visualization, Graphics Interface Conference, pp 163-168.

[6] Koboyashi, T., Misue, K., Shizuki, B. and Tanaka, J. 2006. Information Gathering Support Interface by the Overview Presentation of Web Search Results, Asia-Pacific Symp. on Information Visualisation, vol 60, pp 103-108.

[7] Kules, B. and Shniederman, B. 2004. Categorized graphical overviews for web search results: An exploratory study using U. S. government agencies as a meaningful and stable structure, Proc. 3rd Ann Workshop HCI Research in MIS.

[8] Kules, B., Wilson, M. L. and Shneiderman, B. 2008. From Keyword Search to Exploration: How Result Visualization Aids Discovery on the Web. Technical Report 1516920080208, The Catholic University of America, Washington DC.

[9] Lammarsch, T., Aigner, W., Bertone, A., Miksch, S., Turic, T. and Gartner, J. 2008. A Comparison of Programming Platforms for Interactive Visualization in Web Browser Based Applications, 12th Int. Conf. Information Visualisation, pp 194-199.

[10] Lorigo, L., Pan, B., Hembrooke, H., Joachims, T., Granka, L. and Gay, G. 2006. The influence of task and gender on search and evaluation behaviour using Google, Information Processing and Management: an International Journal, vol 42, no 4, pp 1123-1131.

[11] Nielsen, J. 2006. F-Shaped Pattern for Reading Web Content, Jakob Nielsen's Alertbox, April 2006, online: http://www.useit.com/alertbox/reading_pattern.html last accessed: 10th April 2008.

[12] Reid, J. 2000. A Task-Oriented Non-Interactive Evaluation Methodology for Information Retrieval Systems, Information Retrieval, no 2, pp 115-129.

[13] Reilly, D. and Inkpen, K. 2007. White Rooms and Morphing don't mix: setting and the evaluation of visualization techniques, Proc. of Computer-Human Interaction: Large Displays, pp 111-120.

[14] Reips, U. 2002. Standards for Internet-Based Experimenting, Experimental Psychology, vol 49, no 4, pp 243-256.

[15] Rosenholtz, R., Yuanzhen, L., Mainsfield, J. and Jin, Z. 2005. Feature Congestion a measure of display clutter, Proc. of Conference on Computer-Human Interaction, pp 761-770.

[16] Spink, A., Wolfram, D., Jansen, B. and Saracevic, T. 2001. Searching the web: the public and their queries, J. American Society for Information Science, vol 53, no 3, pp 226-234.

[17] Treharne, K., Pfitzner, D., and Powers, D.M.W. 2008. A Lean Online Approach to Human Factors Research, in proceedings of 1st international conference on PErvasive Technologies Related to Assistive Environments (PETRA), ACM International Conference Proceeding Series; vol 282.

[18] Veerasamy, A. and Heikes, R. 1997. Effectiveness of a graphical display of retrieval results, Special Interest Group Information Retrieval, ACM SIGIR'97, pp 236-245.

[19] Shneiderman, B. and Plaisant, C. 2005. Designing the User Interface: Strategies for Effective Human-Computer Interaction: Forth Edition, Pearson Education Inc.

[20] Grokker http://www.grokker.com

[21] Clusty http://www.clusty.com

[22] Ujiko http://www.ujiko.com

[23] Tianamo http://www.search.tianamo/

[24] EyePlorer http://www.eyeplorer.com/

[25] Rogers, Y., Brignull, H. and Scaife, M. 2002. Designing Dynamic Interactive Visualisations to Support Collaboration and Cognition, IEEE Proceedings of 6th international conference on Information Visualization, pp 39-48.

[26] Mozilla Browser Extension Architecture http://developer.mozilla.org/en/Extensions

[27] Canvas Element Tutorial http://developer.mozilla.org/En/Canvas_tutorial

[28] Zhang, J. 2008. Visualization for Information Retrieval, The Information Retrieval Series, no 23, Springer e-books, Springer-Verlag Berlin Heidelberg.

[29] Morse, E., Lewis, M. and Olsen, K. 2002. Testing Visual Information Retrieval Methodologies Case Study: Comparative Analysis of Textual, Icon, Graphical and Spring Displays, Journal of the American Society for Information Science and Technology, vol 53, no 1, pp 28-40.

[30] Niemelä, M. and Saariluoma, P. 2003. Layout Attributes and Recall, Behaviour & Information Technology, vol 22, no 5, pp 353 – 363.

[31] Börner, K., Chen, C. and Boyack, K. 2003. Visualizing Knowledge Domains, In Blaise and Cronin (Ed), Annual Review of Information Science and Technology, vol 37, iss 1, pp 179-255.

[32] Fekete, J. and Plaisant, C. 1999. Excentric Labeling: Dynamic Neighborhood Labeling for Data Visualization, CHI'99, pp 512-519.

[33] Hartmann, K., Götzelmann, T., Ali, K. and Strothotte, T. 2005. Metrics for Functional and Aesthetic Label Layouts, Lecture Notes in Computer Science, vol 3638, pp 115-126.

[34] Rivadeneira, W. and Bederson, B. 2003. A Study of Search Result Clustering Interfaces: Comparing Textual and Zoomable User Interfaces, University of Maryland HCIL Technical Report HCIL-2003-36. http://hcil.cs.umd.edu/trs/2003-36/2003-36.pdf

[35] Kartoo http://kartoo.com

[36] Hearst, M. 2005. TileBars: Visualization of Term Distribution Information in Full Text Information Access, Proc. ACM SIGCHI Conf on Human Factors in Computing Systems, pp 59-66.

[37] Hoeber, O. and Yang, X. 2006. The Visual Exploration of Web Search Results Using HotMap, 10th International Conference on Information Visualisation, pp 157-165.

[38] Pfitzner, D., Treharne, K. and Powers, D. 2008. User keyword preference: the Nwords and Rwords experiments. Int. J. Internet Protocol Technology, vol 3, no 3, pp 149-158.

[39] Quintura http://www.quintura.com

[40] Bonnel, N. Lemaire, V., Cotarmanac'h, A. and Morin, A. 2006. Effective Organization and Visualization of Web Search Results, Proceedings of 24th IASTED International Multi-Conference Internet and Multimedia Systems and Applications, pp 209-216.

[41] Sebrechts, M.M., Vasilaki, J. Miller, M.S., Cugini, J.V, and Laskowski, S.J. 1999. Visualisation of Search Results: A Comparative Evaluation of Text, 2D, and 3D Interfaces, Proceedings of the 22nd Annual International ACM SIGIR Conference, pp 3-10.

[42] Bonnel, N., Cotarmanac'h, A. and Morin, A. 2005. Meaning Metaphor for Visualizing Search Results, Proceedings of the 9th International Conference on Information Visualization IV'05, pp 467-472.

[43] Rodrigues, J.F., Traina, A., de Oliveira, M.C., and Traina, C. 2007. The Spatial-Perceptual Design Space: a new Comprehension for Data Visualization, Information Visualization, vol 6, pp 261-279.

[44] Skupin, A. and Buttenfield, B. 1997. Spatial Metaphors for Visualizing Information Spaces, Proceedings of Conference on Geographic Information Systems, GIS/LIS'97, pp 116-125.

[45] Perez, C. and de Antonio, A. 2003. Spatialization Algorithms for Text Document Set Visualization, Technical Report CP-03-01, Faculty of Information Technology, Polytechnic University of Madrid, Spain.

[46] Tory, M., Sprague, D.W., Wu, F., So, W. and Munzer, T. 2007. Spatialization Design: Comparing Points and Landscapes, IEEE Transactions on Visualization and Computer Graphics, vol 13, no 6, pp 1262-1269.

[47] SearchMe http://www.searchme.com

[48] Nowell L., France R. and Hix D. 1996. Visualizing search results: some alternatives to query-document similarity, Proc. 19th Ann. Int. ACM SIGIR Conf. Research and Development in Information Retrieval, pp 67-75.

[49] Koshman, S. 2005. Testing User Interaction With a Prototype Visualization-Based Information Retrieval System, Journal of the American Society for Information Science and Technology, vol 56, no 8, pp 824-833.

[50] SearchCrystal http://www.searchcrystal.com/home.html

[51] Spoerri, A. RankSpiral: Toward Enhancing Search Results Visualizations, IEEE Symposium on Information Visualization, pp 18-19.

[52] Havre, S., Hetzler, E., Perrine, K., Jurrus, E., and Miller, N. 2001, Interactive Visualization of Multiple Query Results, Proc. IEEE Symp. Information Visualization, pp 105-112.

[53] TouchGraph http://www.touchgraph.com

[54] Chu, S., Chen, J., Wu, Z., Chu, C.-H.H., and Raghavan, VA. 2007. Treemap-based Result Interface for Search Engine Users, Lecture Notes in Computer Science, vol 4557, no 1, pp 401-410, Springer-Verlag, Berlin.

[55] Walk2Web http://www.walk2web.com/

[56] Apache Lucene Full Text Indexing Library http://lucene.apache.org/java/docs/index.html

[57] Grewal, R.; Jackson, M.; Burden, P. and Wallis, J.; 1999. Novel Interface for Representing Search-Engine Results. Lost in Web Navigation on the Internet: IEEE Informatics Colloquium, pp 1-10.

[58] Carey, M., Kriwaczek, F. and Rüger, S. 2000. A Visualization Interface for Document Searching and Browsing, Proceedings of CIKM 2000 Workshop on New Paradigms in Information Visualization and Manipulation.

[59] Berendonck, C and Jacobs, T 2003. Bubbleworld A New Visual Information Retrieval Technique, Australiasian Symposium on Information Visualisation, in Conferences in Research and Practice in Information Technology, vol 24, pp 47-56.

[60] Cho, E. and Myaeng, S. 2000. Visualization of Retrieval Results using DART, Proceedings of RIAO 2000 Computer Assisted Information Retrieval.